of the parameters that have to do with these methods, it will be closer.

So, I would say we're still in the holding situation, although we're better off than we were in July.

DR. LEE: Marv, you were about to offer some direction.

DR. MEYER: Okay. One of the questions really, in terms of getting some additional data, and I think Dale pointed out quite correctly, to me I would be very convinced if I had five drugs, two bioequivalent and one bio-inequivalent of each one, studied in three labs. That would make the vote easy. Unfortunately, we will have to wait quite a while to get that data, I imagine, if we'll ever get it.

So, I think we have a choice of leaving this guidance that is of questionable, broad application on the books or withdrawing it. And at some point in time, with further research, sponsored by whoever, bring it back. There's nothing wrong with taking something back and then getting more data and bringing it forth again, is there? I mean, that's certainly a viable approach.

I just don't hear right now a convincing set of data to allow this thing to continue to linger out there and generate debate.

DR. LEE: Ajaz?

DR. HUSSAIN: Marv, your suggestion of the difference, we probably would not have the clinical studies to back up the decisions. But if the studies are done that show a difference, that DPK is able to pick up 20 percent, whatever difference -- and we already have some data -- but do that across therapy categories and formulations, would that be helpful to you?

DR. MEYER: It certainly would be helpful. There are other issues, like is the stratum corneum an appropriate sampling compartment, and I don't know that, but I hear experts question that as a place to sample. That would need to be resolved also.

DR. LEE: Dr. Wilkin?

DR. WILKIN: I think it's nice to know that you can pick up different concentrations of an active from the same vehicle. I just remind the group that really the key question is, can you detect differences when the active is at the same concentration but you've got different vehicles. So, it's helpful to be able to see different concentrations in the same vehicle. I would say that that's probably necessary information, but probably not sufficient. I think that's another way of saying what you just said, but I would agree with that.

I think at the end of the day we need to know more about differences in vehicles, some of which are

Q1-Q2, and others which are not Q1-Q2.

DR. LEE: Yes, Art?

DR. KIBBE: A couple of things. The data we saw today was on a gel, which of course is probably the most homogeneous semi-solid we ever use, and it's as close to a solution that we get in a semi-solid. So, if there was going to be a neat system to work on.

Looking at the three items up there, I think issue one, I would have to change viable to possible. I'm not ready to say it's viable.

Issue two, I still think that the two labs got two different answers, even though the labs say they got the same answer. So, I don't think we've gotten to two, or at least we've demonstrated two.

And then I think Marv is right. We need some more studies to get to issue three.

DR. LEE: Kathleen?

DR. LAMBORN: I would suggest that the answer is that it certainly is not a demonstrated method at this point that's sufficient, and that one of the things that should be considered, one of the things I keep hearing around the table is the variety of different types of topical products that current guidance is applying to. And I would suggest that if it is to be re-thought, perhaps withdrawn and at a future date brought back, perhaps a more

focused guidance that would apply to an area that's felt that this technique would be most applicable might be a way to move this back into a procedure. I'm very uncomfortable with this being termed as the method for this full range of techniques. It might be that an incremental approach might help some.

DR. LEE: Jurgen?

DR. VENITZ: Given the history of this, going back over 12 years, and I had the fortune or misfortune of attending the previous meeting a year ago jointly with the Dermatology Committee, my answer to issue one hasn't changed. I don't think this is a viable method. I don't think we can go back and collect the data that we would really need to assess that because that data doesn't just depend on showing bioequivalence or inequivalence in the DPK scenario, but also linking it to the clinical studies. From what I hear you say, and I think the same issue came up last year, for most products we don't really have that endpoint.

So, additional data to assess the technical side of what you're doing right now I don't think would satisfy my issue because my issue is that I can't link what you're testing to clinical endpoints, which presumably we are trying to predict in terms of therapeutic substitution.

DR. LEE: Lemuel.

DR. MOYE: While I don't think the disparate results between the two external labs is the last nail in the coffin of this procedure, I do think it's an important setback. I think that if a procedure is to be viable, it certainly has to be reproducible using the same drug.

Now, the experimental methodology apparently is very complex, perhaps more complex than was initially envisioned. Those inter-experimental methodologic differences are going to have to be worked out, I think, first before we expand the examination and go to different drugs. So, I would say to number one that it is not viable now, and I don't think that we can really address issues two or three until we get the inter-experimental methodology differences worked out.

DR. LEE: Bill?

DR. BARR: I would agree. It seems to me that if we looked at these two studies and found that they agreed, we would all agree that we ought to move ahead with this method. This method has the advantage that it does give us a means of looking at the time course of transport of the drug, which is of course what we do in bioequivalence, and we never usually worry about whether or not at that specific point we can relate that to the clinical efficacy. We usually separate those two and try to state, first of all, we need to know whether or not the

transport to some potentially active site would be the same.

What we see is something that's quite different, and we have one reconciling study which has two people that have been used in it, and perhaps a hypothesis of what these differences are. It seems very clear to me that that next step has to be done to reconcile that before we can go on, and I would suggest that the FDA put some resources into that to perhaps look into that in a little bit more detail, to try to at least resolve that issue before we try to make a judgment.

DR. LEE: I would like to invite the committee members who have not yet spoken to express an opinion if they so choose. Judy?

DR. BOEHLERT: I would agree with the comments that have been made. It's not a viable approach at this time. I'm troubled by the discrepancies in results between Dr. Franz's lab and Dr. Pershing's lab. In my mind that lends to the development of a test that will give you the results that you want, that you can manipulate the test to get the results that you want, and that's not what we want for a regulatory guidance. The area of stripping apparently was important here. So, I can design the area of stripping to get me the result that I want, and that's not appropriate for regulatory guidance. So, we need to do

some more work on how the test is conducted.

I'm also troubled by the fact that we may not have tested all the different types of dosage forms that are out there, creams and ointments and different delivery systems, some of which the drug is in solution, some of which it's not. And would we see the same results if we looked at all of those diverse systems.

DR. LEE: Gloria, you would like to comment?

DR. ANDERSON: I guess I really don't have anything different than what has been said, other than the fact that early on in your presentation you mentioned the fact that it is not known whether or not the uptake at that site -- and I'm a chemist so I won't try to use these biology words -- that that is the only method of uptake. And given the fact that in these studies, both of these studies I think, the methodology involved wiping off the excess of the cream or the ointment or thte gel or whatever it was, and throwing it away without doing I guess a weight balance. Weight balance was mentioned by someone, but it appears to me that that was not an accurate weight balance because if you wipe something off, if you wipe the excess off with a Kimwipe and throw it away, then you don't know how much is on there. It seems to me like that might give some idea of whether or not the uptake is equivalent to the loss from the patch or whatever it's called.

DR. LEE: Okay, so what I've heard this morning is that the situation is far more complex than we envisioned, and it seems to me that the committee is not comfortable to agree with issue number one as stated. So, is it a plausible method but not a viable method.

And issue number two is, does the DPK approach show an appropriate level of between-lab consistency? Based on what we saw this morning, then the answer is no.

And issue number three is -- it's too long. It should not be so difficult or complex? Well, I think the answer is obvious.

Is the committee comfortable with that summation? If so, thank you very much.

That concludes the first session, and let's say that we propose to have a 5-minute break and come back at about 10 after 11:00. Thank you.

(Recess.)

DR. LEE: In the open public hearing, here are the ground rules. Each presenter is going to have 5 minutes to make a presentation, 1 minute to answer questions, if any. The first two address the issue about the derm guidance, and the last few pertain to the IBE for this afternoon.

So, Dr. Spear is at the podium. And, Dr. Spear, are you ready?

DR. SPEAR: Yes, and I'll keep it to 5 minutes.

DR. LEE: Thank you.

DR. SPEAR: Spear Pharmaceuticals has, as you know, supported the study of Franz and Lehman. The FDA sponsored the skin stripping study of Dr. Pershing, and this was a critical step forward in accepting the draft guidance for all dermatologic drugs. Realizing the importance, I felt that it was important to commission Dr. Franz to perform a similar study at another site so that we can really look at this scientifically.

There's no financial connection between DermTech International and Spear Pharma, and the product was sent blinded to Dr. Franz.

The big issue was, is this test rugged? Will the two top places in the country that perform skin stripping report the same results? And we've already discussed this.

Derm products have various sites of action in skin. Skin stripping is really stratum corneum stripping. It really shouldn't be calling it skin stripping.

Antivirals and antifungals act very superficially. Skin stripping theoretically may be the right test, but there's no available data still today -- and I think that's what the committee is wrestling with -- to confirm that skin stripping is predictive of action

below the stratum corneum. For example, anti-acne drugs and corticosteroids. I'm going to keep my comments today regarding tretinoin.

There's still no data on the effect of diseased skin. In dermatology we're dealing with diseased skin states like acne, psoriasis, or eczema where the normal stratum corneum is disturbed. It is a leap of faith to say that how skin stripping behaves on the inner arm of normal skin predicts the effect of drugs in diseased skin, and that's one of the big rubs.

Now, let's look at the two sites here, and we've done this today and I'm going to go very quickly. For a test to be rugged, slight differences in materials or techniques should not really affect the comparative results. Whether or not it's a little bit bigger, a little bit smaller should not really affect that this test is rugged. They really follow the same draft guidance. And we sent Dr. Franz's study to the FDA to review and they actually changed it so the same amount of drug was applied.

Now, comparing Avita gel to Retin-A gel, Dr. Pershing shows lower AUC and lower Cmax, indicating that it absorbs less. Dr. Franz shows it absorbs more. So, the conclusion, my conclusion is that if the two top DPK research sites in the country get contradictory results, the skin stripping methodology is really not adequately

developed. The draft guidance seeks to imply the skin stripping test to all dermatologic drugs. We cannot comment on other classes of derm drugs, but in this example, for anti-acne tretinoin, skin stripping is not rugged.

The draft guidance says, "comparative clinical trials are difficult to perform, highly variable and insensitive." We performed four 400-patient clinical trials on tretinoin products showing bioequivalence to the brand. A skin stripping study today is certainly as difficult to perform, and it seems as highly variable and insensitive.

Some claim the draft guidance must be accepted because generics cannot be proved in any other way. Clinical trials can be done and remain the only confirmatory studies for drugs that act below the stratum corneum.

My conclusion is the FDA seeks to lump all dermatologic drugs into one test. However, there is a movement, and I'm listening today, that they're re-evaluating this position.

Remember, the skin is complex and has multiple sites of action, and we believe that one test does not fit all. We suggest the draft guidance be amended to include, maybe at this point a compromise, only stratum corneum

drugs.

Other DPK tests should be investigated for the deeper action drugs, like the cadaver skin test, with the data that Dr. Franz showed, and not just close your mind and put all your eggs in one basket with skin stripping.

Now, also we've talked here today about clinical relevancy. I'm going to point out two very important points. First, how does Avita penetrate? Avita is promoted as less irritating, so the nice neat little package was to say that clinically it's less effective. But here is a study in the Journal of Pharmaceutical Science, performed by the company who brought out Avita by Penederm with this poly-polymer that they said that it reduces tretinoin penetration, while enhancing epidermal deposition compared to Retin-A. Enhancing epidermal. So, if you're stripping the epidermis, the stratum corneum should have more. Therefore, with skin stripping you should really have more Avita gel with a higher AUC, consistent with Dr. Franz's results.

Let me also point out, in the Journal of the American Academy of Dermatology, the published clinical results of Dr. Lucky, in 215 patients, they showed no difference in total lesion counts at 12 weeks for Avita gel versus Retin-A gel. That's your clinical relevancy there.

Another point that I'd like to make is I went

back in the medical officer review and looked at why the FDA has on there that Avita is less effective than Retin-A. What happened was, there were at two multi-sites, and one of the NDA rules is you must have two independent studies. So, one of their researchers, who was Dr. Jarrett, was dropped from both studies, and he donated 57 percent of the patients and 49 percent of the patients. When his data was included, they were bioequivalent. When his data was dropped out of there, then it showed that Avita was less. Dr. Jarrett was included in Dr. Lucky's publication. So, actually the clinical results show that Avita and Retin-A are actually at 12 weeks the same.

That concludes my comments. Thank you very much.

DR. LEE: Thank you very much. Any questions?

(No response.)

DR. LEE: Thank you.

The next one is Dr. Chris Hendy.

DR. HENDY: Good morning. Thank you for giving me some time to give you a very short presentation.

As Dr. Conner explained, the DPK has been under discussion for some time, and at the last advisory committee meeting, there were some suggestions from the committee that maybe alternative methods may be considered. At Novum, we always do what the FDA tell us to do. We did

decide to take a look at some alternative methods, and I would very briefly like to present some of those to the committee today.

21 C.F.R. 320.24 states clearly: "The following in vivo and in vitro approaches in descending order of accuracy, sensitivity and reproducibility are acceptable for determining the bioavailability or bioequivalence of a drug product." It goes on to list a hierarchy, the number one of those which is "an in-vivo test in humans in which the concentration of the active ingredient or active moiety, and, when appropriate its active metabolites, in whole blood, plasma, serum or other appropriate biological fluid is measured as a function of time."

Using a pharmacodynamic method like the vasoconstrictor assay for the corticosteroids is the third in the hierarchy, and in fact the last acceptable method, fourth, is using comparative clinical endpoint studies.

Many topical products are also available in oral formulations with the same indication as the topical formulation. They also cover a wide range of indications. I've listed some examples there. I'm sure the dermatologists amongst you will be able to give me many more, but that is a whole wide list of indications and different types.

The fact that many topical products are also available as oral formulations means that circulating blood levels must be relevant to the safety and efficacy of the product. Many times the site of action is right next to the blood level, as Dr. Conner pointed out this morning. Bioequivalence of the oral formulation would be evaluated by measuring blood concentrations.

The current draft guidance for industry, the one we've been talking about today, confirms the hierarchical requirement of the Code of Federal Regulations with the following rider. "For topical dermatological drug products, PK measurements in blood, plasma, and urine are usually not feasible to document BE because topical dermatological products generally do not produce measurable concentrations in extracutaneous biological fluids."

Since the development of this guidance and the development of new and more sensitive analytical assays, this statement no longer holds true. The following data are examples of a variety of different topical products where the time course of absorption and elimination of the active moiety can be accurately characterized. In all the examples I'm about to show, the amount and method of drug application is consistent with the product labeling. So, we haven't significantly overdosed to get levels. It's consistent. We haven't left it on for longer than the

product labeling would recommend.

Unfortunately I can't give the drug names because data has been given to me by some of our sponsors. They don't want me to reveal who they are or the actual drug because it is proprietary information. I can tell you it's an antibacterial. This is a two-way crossover study comparing a test and reference formulation. As you can see, the two curves are quite close to each other. However, this product would not pass bioequivalency 80 to 125 percent confidence intervals.

This is another product. This is another antibacterial. Again you can see we can easily measure the concentration in the skin. This is usually a twice-a-day formulation, and you can see this is following a single application, left on for 12 hours.

This product actually is from a full bioequivalency study. This product does meet confidence intervals according to current FDA guidelines. For anyone who had any doubts about the skin acting as a reservoir, this product was actually removed from the surface of the skin at 4 hours, consistent with the product labeling, and as you can see, the Tmax is not until 8 hours.

This is another product, just comparing a small pilot study, again. This is a different route of formulation, but does qualify as topical, and again you can

see we get a nice PK profile.

Many topical products are absorbed to such an extent that the measurement of the active moieties in biological fluids is feasible, as I've demonstrated here with four different products. Many topical products have a site of action such that circulating blood levels are relevant to their efficacy as they're also available in oral or other formulations.

Some topical products are absorbed to such an extent that circulating blood levels could pose potential safety issues. There are several topical products on the market that do have a statement similar to that in their product labeling.

Most topical products have very poor clinical efficacy dose relationships, and I think that's well known.

Clinical efficacy studies are the least sensitive method of determining bio-inequivalent formulations, and our goal must be to make sure that we are not putting bio-inequivalent formulations onto the market.

Evaluating systemic absorption raises the bar for the generic formulation, as it's the most sensitive in determining a bio-inequivalent product than any other of the current methodologies.

And I would suggest that using a pharmacokinetic approach, as demonstrated here, is

consistent with the requirements of 21 C.F.R. 340.21.

DR. LEE: Thank you.

Any questions? Dr. Wilkin?

DR. WILKIN: Well, we had quotes from the C.F.R. but we didn't actually have all of the quotes in the section that I think adds some context to this. The part about the in vivo test in humans, in which "the concentration of the active ingredient or active moiety, and, when appropriate its active metabolites in whole blood, plasma, and serum" -- and that's the part that you quoted. But the sentence that follows that I think is also important to the understanding. It says: "This approach is particularly applicable to dosage forms intended to deliver the active moiety to the bloodstream for systemic distribution within the body."

Then if you go on further down in the same section, it speaks to the clinical studies and it says: "This approach may be considered sufficiently accurate for determining the bioavailability or bioequivalence of dosage forms intended to deliver the active moiety locally." Topical preparations to the skin. It gives some other examples.

Just to clarify that the C.F.R. I think has a somewhat slightly different view on that.

On the other hand, I would say this is an

exciting thing to think about, that the limits of detection with newer technologies have gotten to the level where now we could look at blood AUCs.

I think we have to remember, though, that it's how the drug is distributed to the active site in the skin, which is a very heterogeneous organ. There are a lot of different active sites, and will the vehicle send the active down the follicle, or will it go through the epidermis between the follicle? So, it actually could end up in the blood at the same rate in the end, but it might get there by different pathways. On the one hand, it might bypass the critical place in the skin where we're really ultimately interested in rate and extent, but nonetheless it's certainly an interesting thought.

DR. LEE: One minute to answer.

DR. HENDY: I absolutely agree with Dr. Wilkin's comment. But obviously a number of the products we've put up here do not act in the stratum corneum. We know that from their pharmacology, and they are going to be working a lot closer to circulating blood levels, and one would assume that there is some kind of homogeneous area there. But I'm not a dermatologist so I really can't go on further than that. But I do think it's a methodology that maybe we should be looking at as an alternative to some of those that are being suggested.

DR. LEE: Thank you very much.

Okay, we now move into the IBE positions. Dr. Sondhi?

DR. SONDHI: This paper has been submitted for publication, so that's why I think you didn't get copies of these in the handout.

What we are trying to show is that you can get a probability distribution of the bioequivalent metric, and I just wanted to show that.

The metric was defined by Hyslop as follows. You see that on the viewgraph there. P equals mu T minus mu R squared, et cetera, where the mu's are the means of the pharmacokinetic parameter for the test and reference products. The sigma's are the test and reference variances within subject, and sigma i squared is the variance of the difference in the means.

For sample values of the test and reference means and variances, and the X's and S's that I've shown there, for those sample values you can get an estimate of phi, which is now a random variable.

Now, the problem is, of course, to find the 95th percentile of the probability distribution of the phi hat and accept bioequivalence if the value is below the FDA-specified value.

Hyslop, et al. found the upper 95 percent

confidence interval of a linearized version of this metric. Instead, what we are proposing is that it's also possible to, in fact, get the entire probability distribution function, or an estimate of that probability distribution function, whose interval then gives us the cumulative distribution. If the 95 percent point of the cumulative distribution is below the FDA-defined value, we accept bioequivalence.

Now, the probability distribution of phi hat can be determined if the joint distribution of all those variables that I've shown there is known, but of course, in general, it would be a very formidable task. However, under the usual assumptions of statistical independence of these variables, the computation is quite feasible. And that's the purpose of this paper, to show that it's quite feasible.

I might say, of course, you get an approximation to the probability distribution because we substitute sample values of the means and variances, since the actual values, of course, are not known.

Just to make the notation simpler, I just gave names to these parameters. Xt minus Xr is Y, and Si squared is Z and so on. So, if you write it in this way, all we can say on the bottom is that the metric phi is just this ratio of G over V minus 1.5.

I obviously won't give you the derivation of this probability distribution, but just tell you the steps involved. If you assume the Xt and Xr to be independent, then you need a formula to find the sum of the square of the difference Xt minus Xr. And that's a known formula which one can use.

Then you can compute the PDF of W, which is the sum of two random variables, independent random variables, and that we know how to do.

Then the distribution of G we do the same way because it's the sum of two other independent variables.

Then the ratio G over V, we need a formula for finding the distribution of the ratio of two independent variables, and that's fairly well known.

With these few steps, one can then get the probability distribution of the metric phi.

I've written a program for this, and once the program is written it, of course, runs in a few seconds, so I'll just give you two graphs showing the examples of using this method.

Here is a comparison of results by the two methods, Hyslop's and the one that we are proposing here. I'm showing just two cases of situations where the pass-fail is right at the boundary. In other words, they're very sensitive measurements. So, you can see that in all

of these cases the pass-fail was exactly the same for Hyslop's and with us. Very rarely do we find any difference in the decision.

This is not a very good graph because the action is taking place only on the first inch or so of it, but this is a plot of the entire cumulative distribution for a particular set of parameters.

DR. LEE: Questions for Dr. Sondhi?

(No response.)

DR. LEE: If not, thank you very much.

Professor Endrenyi?

DR. ENDRENYI: First, I'm grateful for the opportunity to be able to be here. CDER has known that I haven't always agreed, and I think it is very gracious. But I still don't agree.

The first suggestion is that individual bioequivlaence in practice has unfavorable properties, and I underline "in practice." The acceptance or rejection of individual bioequivalence can be due to random chance alone.

To demonstrate, in the model of individual bioequivalence, as you are going to see this afternoon, an important term is the difference of within-subject variance of the test and reference products.

Another term is the difference between means,

and the question is how these two terms play against each other. That's called mean variance tradeoff.

Generally speaking, there is reward if the variance within-subject variation of the test formulation is smaller than the within-subject variation of the reference formulation, under this condition. Then in that case the test formulation is better. So, in contrast, that would be a penalty if the test formulation has a higher variation than the reference formulation. This is what the model says.

But in practice, when it comes to estimated variations, if the true variations of the two formulations are identical, then it makes sense that in practice the test formulation and variation can be higher, estimated variation, or lower than that of the reference formulation and actually that these conditions can occur with equal probability.

This is what this slide demonstrates for actual data which the FDA collected by '99. By now they have additional data. You see here that the reward condition and penalty condition occur with about equal frequency. Furthermore, large rewards and large penalties also can occur with fairly high frequency and usually with equal frequency, or similar frequency.

This follows, then, theoretical considerations,

and the consequence is that the acceptance and rejection of individual bioequivalence can be due to random chances alone.

Turning to higher variable drugs, which is one of the two drug classes for which replicate designs are recommended, the analysis of trials that scaled individual bioequivlanece, reference scaled individual bioequivalence be used, but we contend that scaled average bioequivalence is much more effective for the purpose.

Now, the next slide would demonstrate this, but I shall turn to it only if there is time.

Again, the following slide demonstrates the next statement, namely this, which has to do with the proposed ratio of geometric means, the GMR. In the guidance, GMR for individual bioequivalence of 1.25 is recommended, and in the demonstrations we show the scaled average bioequivalence and this constraint is workable. Doesn't change much the character of the test.

On the other hand, the test of the scaled individual bioequivalence and the constraint dramatically changes the individual bioequivalence test. It simply becomes not an IBE test but becomes a GMR test, a test of the geometric ratios.

The same condition can be expected if one constrains the ABE test down to 1.15 or 1.10. It becomes

probably -- because we haven't done these studies -- but probably the consequence is that we would have a GMR test rather than an average bioequivlanece test.

So, we think that the imposition of a very narrow constraint would change the character of the test and would be probably counterproductive. Therefore, we conclude that the acceptance or rejection of individual bioequivlaence can be due to random chance alone, and therefore it's not really a good procedure, in our opinion. Scaled average bioequivalence is much more effective than scaled individual bioequivalence for assessing highly variable drugs. And moderate constraint could be workable for scaled average bioequivalence, not for scaled individual bioequivalence. A strong constraint would probably be counterproductive.

I have additional slides, but no time. Thank you.

DR. LEE: Thank you very much. Any questions for Professor Endrenyi?

DR. MEYER: Laszlo, could you amplify point three a little more for me? Why 1.15 would be probably counterproductive? Because that speaks, it seems to me, to one of the issues of confidence in the FDA's decision and 25 percent difference is larger than we're used to.

DR. ENDRENYI: In this case, consider the

scaled individual bioequivalence curve, which is this curve. And consider the geometric limit alone, which is this. What you have is the acceptance of tests as you vary the true ratio of the geometric means. That's further separations.

First of all, you notice that the individual bioequivlaence curve is a very permissive. It permits large deviations.

The general rule is that when two criteria are joint -- in this case, they're individual bioequivalence, and they're GMR criteria -- then the joint criterion has acceptances which are lower than either in the separate criterion. That makes sense.

Now, in this case, when the GMR criterion is so much tighter than the IBE criterion, then the joint criterion actually draws close to the GMR criterion. So, it becomes a GMR criterion rather than an IBE criterion.

The same thing would happen when the ABE, average bioequivalence. When the GMR criterion moves to the left because of time constraint, it's well to the left of the ABE criterion, and the GMR criterion would dominate.

DR. MEYER: If you had scaled average bioequivalence, and you had the 1.15 GMR, wouldn't you still have potentially the larger confidence intervals that would pass?

DR. ENDRENYI: Without the GMR, yes. With the GMR you would have the GMR criterion.

DR. MEYER: So, you couldn't have confidence limits that were beyond 1.25.

DR. ENDRENYI: That's right. Essentially you would have the Canadian Cmax criterion.

DR. LEE: Okay, on that note, thank you.

Mr. Charles Bon?

MR. BON: Actually I'm going to address in part some of what Laszlo said. I want to thank you for the opportunity to address the committee.

I wanted to start with just a brief discussion of what the individual bioequivalence criterion is based on. It starts with the ratio of the expected square of changing a patient from the reference product to a generic test product to the expected squared ratio of that same patient taking the reference product on two different occasions, and then we place some limits on that.

In the development of the criterion, it's an aggregate criterion. It was elegantly developed through mathematical and statistical considerations. In the criterion, you've seen there's a difference in means, a difference in variances, and a subject-by-formulation term, which really talks about the consistency of the test to reference response in the subject studied. In the case of

highly variable drugs, it's scaled by the reference variance.

However, what happened to the criterion was just what Marvin had said. There were things that we weren't used to. One of the things was to allow the test-to-reference ratio of means to go outside of the .80 to 1.25, so this was really added without justification. I see in what you're going to be asked to look at this afternoon is to further restrict this as well as to put a restriction on the subject-by-formulation term.

Proposing that the restriction on these individual terms in the aggregate criterion is not supported by the mathematics and the development of the theory. It's not supported by any clinical or good scientific considerations, and it has very undesirable consequences.

I'm going to show you the results of a small pilot crossover study on an immediate-release coated oral tablet that the FDA had approved in the early 1980s. This was just a single one-tablet fasted dose in generic versus brand.

In this we found that the log AUC's were comparable in terms of both the observed ratio and it actually meets the .8 to 1.25 confidence interval. Log Cmax was well outside certainly on the confidence intervals

and on the individual ratio, and yet the Tmax's were very similar.

I'll show you a couple of examples here of some of what I call the well-behaved subjects. I'll just show you a couple of these subjects, where the test is in red and the other color is the brand. But we had two subjects out of the 10 that gave very low profiles on the brand, even though their profiles on the generic product were quite consistent with those of the other subjects. In fact, we saw profiles on the brand that didn't really look like it was an immediate-release product.

In going back and looking at the formulation, I'm told by the formulators that there's this coating on the tablet in the reference product which is old technology and is a very poor coating, and in dissolution there were problems with certain units of the reference product.

Now, here is actually the test-to-reference ratios and I've highlighted two problems. Here are 3.9 and 3.2 for test-to-reference ratios on Cmax. We actually had a couple of other high ratios which may be partial problems with the reference.

But I'm going to use this example with some assumptions. I did a simulation of 100,000 replicated trials, assuming that a good test-to-reference ratio that occurs in 80 percent of the brand tablets with this

particular generic product would be an acceptable ratio of 1.05. But 20 percent of the brand tablets would not release in an immediate-release fashion and give you this lower Cmax resulting in an expected test-to-reference ratio of 3.5.

Consistent with what we saw, the generic product was actually on an inter-subject basis less variable than the reference, but under the assumptions for the simulation, just to illustrate my point, I assumed a 20 percent within-subject within-product CV for the generic, 30 percent for the brand, and I did a replicated study in 30 subjects.

Less than 30 percent of the time the test-to-reference ratio fell within .8 to 1.25. This is just the geometric mean ratio, which immediately, regardless of what else was happening with the aggregate criterion, this product would be deemed to be bio-inequivalent by individual bioequivalence.

The only recourse that the generic company has is to actually make a bad generic product, and a bad generic product that falls somewhere in between the good ratio of 1.05 and this ratio of 3.05 so that they can overcome the rather arbitrary restraints placed on the difference in means or on the geometric mean ratio. This is one of the side effects of placing constraints on

something that is really a good aggregate criteria.

DR. LEE: Thank you very much. Questions?

(No response.)

DR. LEE: And the next one is Mario Tanguay.

DR. TANGUAY: First I would like to thank the committee for this opportunity to present on behalf of the GPhA and MDS Pharma Services some comments on the individual bioequivalence approach.

I would also like to thank the GPhA Science Committee and the CRO Biopharmaceutic Committee for their collaboration, as well as my colleagues from MDS Pharma Services.

The IBE approach offers some advantages compared with the average bioequivalence approach. When we administer the same formulation twice in the bioequivalence study, it allows one to better differentiate the variability associated with each formulation. Contrary to the average bioequivalence approach, the IBE approach takes advantage of the fully replicate design.

The IBE approach may also be advantageous from an ethical point of view, since a smaller number of subjects is required for highly variable drugs. Due to the internal scaling component of the current IBE approach, the widening of the goalposts will depend on the variability of the reference product.

However, there might also be some disadvantages associated with the IBE approach. The main one is that there is some uncertainty regarding the switchability assessment, or, in other words, the subject-time-form interaction.

Bioequivalence studies are designed to compare the relative rate and extent of bioavailability of two formulations of the same active ingredient based on Cmax and AUC calculations. These studies are not primarily designed to rapidly assess switchability.

If a subject-time-form interaction is seen in a study, there is no way to determine clearly the reason for this observation. It is not clear if this could be due to the presence of an outlier, for example, or to a subset of subjects, or if this could be due only to chance. It is also possible that different results with regards to switchability would have been observed if the drug products would be administered more often.

In addition, when the IBE approach is used, it is highly recommended that subjects from a heterogeneous population be enrolled, meaning that people from different age, gender, race and so on should be enrolled. However, this may not be helpful, as these studies are again not designed up front to evaluate differences in pharmacokinetics based on demographic factors. Therefore,

even if the subject-time-form interaction is observed, this will need to be proven further by a properly designed study, which would raise other questions.

In the clinical research area, it has been proven many times that conclusion from posteriori analyses were proven to be wrong when verified in properly designed prospective studies. There are many examples of this situation in cardiovascular pharmacology or infectious disease, for example, and these lessons should apply to switchability measurement in bioequivalence studies.

In conclusion, the bioequivalence of two formulations of highly variable drug will be better assessed by giving the same formulation more than once to the same subjects. The IBE approach can then be useful for highly variable drug products. However, there is some uncertainty regarding conclusions that could be drawn from a switchability assessment.

Thank you.

DR. LEE: Thank you very much. Any questions?

DR. SHARGEL: Just a quick comment on the point that you said, ethics. It is a reduction of blood samples in IBE, but you do have more exposure to the drug by the individual subject. You have twice the exposure, so I think that can be considered as well.

DR. LEE: Thank you, Leon.

Dr. Midha?

DR. MIDHA: I'm grateful to the committee for the opportunity to speak to you today. I'm here on behalf of PharmaLytics, which is a nonprofit institute of the University of Saskatchewan.

You have already heard some very good comments. The important consideration is that highly variable drugs or drug products are safe drugs with flat dose-response curves or shallow dose-response curves. That means otherwise they wouldn't have gotten on the market. So, you're dealing with drugs which are safe.

A drug with ANOVA-CV, an average bioequivalence of 30 percent, has been defined as highly variable. If the drug product has the ANOVA-CV in a two-treatment crossover design, it is then considered to be a highly variable drug product, so differentiate between drug and drug product.

The problem with highly variable drug products are that you need a very large number of subjects in order to meet the average bioequivalence criteria preset, which has confidence limits of .80 to 1.25 percent. You need a very large number of subjects, and people have calculated from anywhere 60-fold to over 100 subjects.

I'm just going to show you an example. Chlorpromazine is an example of a highly variable drug. We had shown 10 years ago or 11 years ago that it had an

average ANOVA-CV, an average bioequivalence of 34 percent for AUC and 43 percent for Cmax. The test product in the study was given once and the reference product from the same lot was replicated.

The next slide shows the results from that study, which have been published. You're looking at the ANOVA-CV's of 34 percent and 43 percent. What you observe here, that when test is compared to reference, it meets the criteria for AUC. When reference is compared to same lot, it meets the criteria. But when you look at the Cmax, with ANOVA-CV of 43 percent, test compared to reference does not meet the criteria because the upper bound is 1.26 percent, above 1.25.

But look at reference to reference. Here now the criteria is violated to the extent that ANOVA-CV takes it other than the geometric mean ratio of 115 percent -- and that's what Marv's question, we are trying to fix. It is now 136 percent. But this product has been on the market and has been utilized for over 40 years. So, it's clear that two samples from the same lot of the reference product were not found to be bioequivalent with each other because, one, the ANOVA-CV was large and the point estimate for reference to reference was 115 percent, a comment Professor Endrenyi made earlier.

These data and the data which we have done

research on demonstrated that the reference formulation was a good quality product, but it was the drug which was highly variable, and the drug has been on the market for 40 years.

Under the new recommendation of October 2000 guidance, when stated a priori, after due consideration with the agency, scaled IBE based on replicate design may be allowed for a highly variable drug and drug product. This in our opinion is a reasonable approach. We were one of the first research groups to make a recommendation to go using replicate design, do average bioequivalence scale, a case which Dr. Endrenyi made again.

But in absence of that, at present the guidance has a very reasonable proposal, and I believe for the trial period it ought to be maintained based on the fact that these drugs are safe and we do not wish to do undue human experimentation when it is not needed.

The use of scaling in a highly variable drug, because you are scaling to the reference variability for the type of variability which already exists in the marketplace, permits the assessment of bioequivalence to be performed with a reasonable number of subjects. Yes, there are replicated measures but they are a reasonable number of subjects without compromising either the consumer risk or the producer risk.

An additional advantage which you heard from the previous speaker is that IBE in a replicate design or average bioequivalence based on replicate design would allow you to look at the pharmaceutical quality of the product. With all the advancement made in pharmaceutical sciences, you would like to see the generic formulations continue to come which have got reduced variability.

The constraint that the GMR must fall within 80 to 125 percent for scaled IBE is reasonable and should be maintained.

Increasing the constraint on GMR in IBE to less than 125 percent is actually defeating the very purpose. We are going back a decade.

Our plea is we should not modify the recommendation in the guidance until scientific evaluation of the scaled metrics are completed.

And I thank you for your attention.

DR. LEE: Thank you very much, Kam.

Any questions? Yes.

DR. MEYER: Kam, you're kind of pleading for more data and more evaluation of existing data. In February of 1998, you and Jerry Skelly and Laszlo and Gordy Amidon published a paper entitled "IBE: Attractive in Principle, Difficult in Practice." You were pleading for more data. It's been 2.5 years now. Surely we have more

data. Can't we either decide one way or the other?

DR. MIDHA: No, I'm not asking for highly variable drugs or drug products to have more data. That paper was written in light of the fact that there was a strong move afoot to apply IBE all across, and that's why the plea was, and the plea continues to be.

I would also go, Marv, and make a plea that we ought to also investigate in replicated design studies average scaled bioequivalence because if we are not prepared to widen the bioequivalence limits based on the class of drugs, I think that may be another reasonable approach. But in view of the fact that we don't have average bioequivalence scaled from replicate design in the guidance, the only step forward is the step which we have taken in the October 2000 guidance.

DR. LEE: Dr. Lesko?

DR. LESKO: Thanks.

Kam, on the framework for your remarks today, you indicated that the problem was highly variable drugs or drug products. My understanding of the framework for this problem is that we have a generic product that approximates a GMR of approximately 1, but in order to meet the 80 to 125, we need a large number of subjects. So, it seems to me reasonable to scale in the context of approximating 1 as the ratio, but what you're basically concluding is that

it's okay to have a 25 percent increase in bioavailability of generic product and then on top of that go ahead and scale.

I guess I'm wondering why you think constraining the mean to 15 percent, which is approximating what would be allowable under the current standard for average bioequivalence, would interfere with the ability to scale bioequivalence limits to allow for lesser subjects for a highly variable drug.

DR. MIDHA: Larry, I probably have not followed your question, and I think I'm going to spend some time discussing with you. But if I have followed it correctly, the reason is that when you take the constraint down to 115 percent, essentially that becomes the determinant step, not the limits. So, the result is that as Laszlo could not show that, that the GMR becomes the determinant in terms of declaring bioequivalence, then whether it is the limits of average, or in the case of IBE the limit. So, unless you want to take that determination -- and I don't want to take too much time of the committee, but clearly that is the crucial issue which we are dealing with.

In the case of highly variable drugs, you know, and we have had many discussions on it, they are safe drugs, otherwise they wouldn't get on the market. And the fact is, we are trying to force the GMR, asking 90 percent

of the time the values are going to exist in that. When reference-to-reference from the same lot can show you those kind of variability. It's already in utilization for over 40 years. So, that's my plea to you and the people, those who are going to consider it.

DR. LEE: Okay, I realize that there are quite a few questions to be posed, but in the interest of time, I'm going to close this morning's session. We have an afternoon devoted to individual bioequivalence, and I saw that 4:30 is the time of adjournment. In order to keep to that I'd like to suggest that we come back here at about 1 o'clock.

Thank you.

(Whereupon, at 12:06 p.m., the committee was recessed, to reconvene at 1:00 p.m., this same day.)